Raising detectability of additional data in a media signal having few frequency components

TECHNICAL FIELD

The present invention generally relates to the field of providing additional data in a media signal and more particularly to methods, devices, a signal and an information storage medium related to embedding of additional data in a media signal.

DESCRIPTION OF RELATED ART

With the evolution of the Internet it is possible to access or retrieve a virtually limitless amount of informational content. Content can then be provided by different content providers in the form of media signals of varying shapes and forms. Media signals can for instance be provided as audio signals, in either compressed or uncompressed form, image signals in compressed or uncompressed form as well as video signals in compressed or uncompressed form. In order to inhibit that media content is unlawfully obtained by persons not entitled to it or that illegal copies of content are being made, there is a need for content owners to protect their content. In order to do this they often need to provide additional information in the media signals. Additional information can also be provided for other reasons, like for instance for providing text in relation to a piece of audio (e.g., lyrics).

One field of use where additional data is provided in media signals is in the field of Digital Rights Management (DRM), where additional data in the form of watermarks are used to indicate the origin of media content and possibly of user in order to inhibit unlawful tampering of the media content.

The possibility of correct and effective watermark detection depends heavily on the method used for embedding the data into the host signal and on properties of this signal. One frequently used type of watermark embedding is the so-called multiplicative watermarking, where the media signal to be watermarked is multiplied with the watermark in question. On the other hand, normally a media signal has a lot of different frequency components, whereas sometimes it can have few such components. When the components are few it can be hard to detect a watermark that has been embedded using multiplicative watermarking.

International patent application WO-A-02/15587 describes how additional data, like a watermark, is added to a media signal. The signal is here described in relation to a sine wave. A binary code is added to the signal in a high frequency band through either adding noise or not adding noise in this high frequency band. Upon detection, the sequence of digits (i.e., zeroes and ones) obtained represents (a coded version of) the watermark information. The document thus describes a technique for additive watermarking, which is not applicable in a multiplicative watermarking environment. Besides, since the additional information is only provided in a high frequency band, which can easily be filtered away using a simple low-pass filter, it is fragile and therefore not suitable when robustness is an important condition.

In a more robust, multiplicative watermarking scheme, a plurality of circular shifted chip sequences of real numbers is multiplied with a properly scaled version of the media signal and added back to the original media signal. Upon detection, the distances between the diverse correlation peaks carry (a coded version of) the watermark information. If the host signal contains few frequency components, the correlation will be weak. There is thus a need for enabling a higher level of detectability for additional data that has to be embedded in a media signal with few frequency components using a multiplicative embedding technique.

SUMMARY OF THE INVENTION

It is thus an object of the present invention to provide multiplicative embedding of additional data in a media signal that is more robust (i.e., has a higher level of detectability of the additional data), especially in sections of the media signal that have few frequency components.

According to a first aspect of the present invention, this objective is achieved by a method of embedding additional data in a media signal comprising the steps of: obtaining a media signal,

mixing at least one section of said media signal with a noise signal for providing a modified media signal, and

combining said additional data with said modified media signal for providing a first host modifying media signal.

According to a second aspect of the present invention, this objective is also achieved by a device for embedding additional data in a media signal comprising:

a first adding unit for mixing at least one section of said media signal with a noise signal in order to provide a modified media signal, and

a combiner unit for combining said additional data with said modified media signal for providing a first host modifying media signal.

According to a third aspect of the present invention, this objective is furthermore achieved by a media signal comprising:

at least one section of modified media signal comprising media signal mixed with a noise signal, where additional data has been combined with this modified media signal.

According to a fourth aspect of the present invention, this objective is also achieved by an information storage medium comprising:

a media signal including at least one section with modified media signal comprising:

media signal mixed with a noise signal,

where additional data has been combined with this modified media signal.

The present invention is furthermore directed towards providing a technique for (automatically) switching between the media signal and a modified version of the media signal in order to selectively enhance the detectability of multiplicatively embedded information to this new host signal.

According to a fifth aspect of the present invention, this objective is achieved by a method of embedding additional data in a media signal comprising the steps of:

obtaining a media signal,

analysing the media signal,

mixing at least one section of said media signal with a noise signal for providing a modified media signal, and

combining, for different sections of the media signal, said additional data with said modified media signal for providing a first host modifying media signal or with said media signal in dependence of the analysis.

According to a sixth aspect of the present invention, this objective is also achieved by a device for embedding additional data in a media signal comprising:

a first adding unit for mixing at least one section of said media signal with a noise signal in order to provide a modified media signal,

a combiner unit for combining said additional data with said modified media signal for providing a first host modifying media signal or with said media signal, and

an analysing unit arranged to analyse said media signal and control, for different sections of said media signal, the provision of said media signal mixed with noise or said media signal to the combiner unit in dependence of the analysis.

Claims 2 and 16 are directed towards performing the combining using multiplication.

Claims 5 and 17 are directed towards shaping the noise signal based on a model of human perception. This has the advantage of making sure that the added noise is not perceptible.

Claims 6 and 18 are directed towards shaping also the modified media signal that is combined with said additional data with a signal shaping function based on a model of human perception. This has the advantage of making sure that both the added noise and the embedded watermark are not perceptible.

Claims 8, 9, 10, 20, 21 and 22 are directed towards scaling the added noise, adding the media signal to the modified media signal that is combined with said additional data and adding the unscaled noise signal to the media signal that is combined with the additional data. This has the advantage of providing a more predictable control mechanism for the embedding of additional data.

Claims 12 and 23 are directed towards analysing the media signal and combining the additional data with sections of the media signal or the media signal mixed with noise in dependence of the analysis.

The present invention has the advantage of providing better detectability of additional data when it is embedded in a media signal having few frequency components, e.g. highly tonal signals like excerpts of pitch-pipe or harpsichord. With the invention it is for instance possible to embed a more easily detectable watermark in a modified media signal compared with an ordinary media signal having these properties. Because of this higher level of detectability the additional data remains detectable even if the quality of the media signal is degraded, i.e. the probability of a correct detection has increased. It is then easier to perform for instance forensic tracking of a processed media signal.

The general idea behind the invention is thus to mix a media signal with a noise signal and combine the additional data with the media signal that has been modified in this way.

These and other aspects of the invention will be apparent from and elucidated with reference to the embodiments described hereinafter.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will now be explained in more detail in relation to the enclosed drawings, where

Fig 1 shows a block schematic of a device for embedding a watermark in a modified media signal according to a first embodiment of the invention,

Fig. 2 shows a block schematic of a first variation of a combiner unit that can be used in the device in fig. 1.

Fig. 3 shows a block schematic of a second variation of a combiner unit that can be used in the device in fig. 1,

Fig. 4 shows a block schematic of a device for embedding a watermark in a modified media signal according to a second embodiment of the invention,

Fig. 5 shows a block schematic of a device for embedding a watermark in a modified media signal according to a third embodiment of the invention,

Fig. 6 shows a flow chart of a method of embedding a watermark in a modified media signal according to the third embodiment of the invention,

Fig. 7 shows a block schematic of a device for embedding a watermark in a modified media signal according to a fourth embodiment of the invention,

Fig. 8 shows a block schematic of a device for switching between embedding of a watermark in an original media signal or a modified media signal according to the invention, and

Fig. 9 shows an information storage medium in the form of CD disc having a media signal according to the invention stored on it.

DETAILED DESCRIPTION OF EMBODIMENTS

The present invention relates to the field of providing additional data in media signals having a sparse frequency content in at least parts of the signal. In the field of audio such signals can include the sound from instruments like harpsichord and pitch pipe. The invention is however not limited to audio but can be applied on other media signals like for instance video or digital images. The additional data is preferably provided in the form of a watermark. It should however be realised that the invention is not limited to watermarks, but the additional data can be any additional data that needs to be detected in a media signal, like for instance additional text in relation to a song.

Fig. 1 shows a block schematic of a device 10 for embedding additional data in a media signal having sparse frequency content according to a first embodiment of the

invention. For this reason the device 10 includes a first adding unit 12, which first adding unit 12 receives the media signal x and adds a noise signal n to this media signal in order to provide a modified media signal x + n. The media signal x is in these circumstances often referred to as the host signal. The modified host signal x + n is then supplied to a watermark combiner unit 14, which combines the additional data in the form of a watermark w in the modified host signal x + n to provide a first host modifying signal m_w at its output. Finally, in a second adding unit 36, the first host modifying signal m_w is added back to the modified host signal x+n (or the host signal x) to provide an output media signal y with said additional data. The combiner unit 14 shown here is a filter that applies the watermark w in the form of suitably selected filter coefficients. The combiner unit 14 is thus a multiplicative unit that modifies the modified host signal x + n through multiplying it with the watermark. Because the modified signal contains more frequency components than the original signal, the watermark is easier to detect. The noise signal n is here an additional watermark carrier, so that both the noise signal and the host signal carry the watermark.

However, also signals that have many different frequency components may benefit from this type of embedding, especially by insertion of noise shaping in the higher frequency range. This will not significantly improve robustness of the watermark, but for unprocessed watermarked audio it may yield significantly better detection reliabilities.

Fig. 2 shows a first variation of the combiner unit 14 according to the invention, which works in the frequency domain. The combiner unit therefore includes a discrete Fourier transform unit 16 which receives the modified host signal x + n and transforms it to the frequency domain. The transformed modified host signal is then provided to a multiplying unit 18, which multiplies the transformed modified host signal with a watermark w. The watermark w is here a frequency domain watermark. The watermarked transformed modified host signal is then provided to an inverse Fourier transform unit 20, which transforms the watermarked transformed modified host signal back into the time domain and supplies it to a multiplying unit 22. The multiplying unit 22 also receives the results from a graceful raising/decaying on/off switching function. In order to provide this switching the modified host signal x + n is therefore supplied to a unit 24, which uses a temporal gain function G. The output of the multiplying unit 22 is then provided to a scaling unit 26, which scales the multiplied signal with a scaling parameter α. This multiplied and scaled signal is then provided to the second adding unit 36, which also receives the modified host signal and adds these signals together to form the output signal y, which is the watermarked host signal. More details about embedding of watermarks according to this

principle is described in the document "Robust, multi-functional and high-quality audio watermarking technology", by Michiel van der Veen, Fons Breukers, Jaap Haitsma, Ton Kalker, Aweke Negash Lemma and Werner Oomen in The Proceedings of the 110-th AES Convention, Amsterdam, The Netherlands, May 2001, which is herein incorporated by reference.

The above described frequency domain combiner unit can be modified in many ways. It is for instance possible to remove the branch including the amplifying unit and also to remove the scaling unit, although this would degrade the signal quality.

Fig. 3 shows another variation of a combiner unit that works in the time domain. The combiner unit 14 includes a bandpass filter 30, which filters the modified host signal x + n and provides the filtered signal to a multiplying unit 32, which also receives a watermark w and multiplies the watermark w with the filtered modified host signal x + n. The output of the multiplying unit 32 is connected to a scaling unit 34, which scales the watermarked signal with a scaling parameter α and provides it to the second adding unit 36, which also receives the modified host signal x + n. The output of the second adding unit 36 is then the watermarked host signal y. The scaling unit 34 is also here not strictly necessary for providing a watermarked signal. The watermark w is here a time domain watermark. More detail about this watermarking technique can be found in the document, "A temporal domain audio watermarking technique", by Aweke Negash Lemma, Javier Aprea, Werner Oomen and Leon van de Kerkhof, IEEE Transactions on Signal Processing, April 2003, Vol. 51, page 1088-1097, which is herein incorporated by reference.

The above described combiner units are just examples of multiplicative combiner units than can be used in the present invention. It should be realised that many other types of multiplicative combiner units can be used instead.

The thus described watermarking technique shown in fig. 1 can be improved in that a model of human perception can be used for shaping the noise signal for reducing the perceptible distortion. The model used depends on the type of signal. In case the media signal is an audio signal the model is a psychoacoustic model of the human hearing system and in case a pure image is used a psycho-visual model of the human visual system is used.

A block schematic of a device for performing embedding of a watermark into a media signal according to a second embodiment of the invention is shown in fig. 4. The device in fig. 4 basically includes the same components as the device in fig. 1. There is one difference though and that is that the device 10 further includes a first signal shaping unit 40 in the form of a masking filter and a filter control unit 38. The filter control unit 38 receives

the host signal x, analyses this signal using a psycho-acoustic model of the human auditory system P. The unit 38 uses the results from the analysis for choosing filter coefficients of the filter 40. The filter 40, which receives the noise signal n, shapes the noise using a first signal shaping function M1 so that a shaped noise signal n_s is obtained. This shaped noise signal n_s is then provided to the first adding unit 12 for mixing with the host signal x. Thereafter embedding of a watermark is performed in the above described way in the watermark combiner unit 14. The filter 40 shapes the noise signal so that it is perceptibly masked by the host signal x. If the media signal were an image the model would be a psycho-visual model of the human visual system instead.

It is possible to further vary the device according to the invention by also including a second signal shaping unit using a signal shaping function M2, which is also based on information from the filter control unit 38. A device according to this third embodiment is shown in a block schematic in fig. 5. The functioning of the device in fig. 5 will now be described also in relation to fig. 6, which shows a flowchart of a method according to this third embodiment. The noise adding is in this embodiment the same as the noise adding in fig. 4. The only difference here is that the device 10 includes a second noise shaping unit 44. First a host signal x is obtained, step 48, for instance by fetching it from a memory where it is stored. The noise signal n is provided, step 50, for instance from a noise generating unit. Thereafter the noise signal n is shaped using the first noise shaping function M1 in the filter 40 for obtaining the shaped noise signal n_s, step 52. The shaped noise signal n_s is then added to or mixed with the host signal x by the first adding unit 12 in order to provide the modified host signal $x + n_s$, step 54. The combiner unit 14, which can for example be only a filter or one of the units shown in fig. 2 or 3, receives the modified host signal $x + n_s$ and combines the watermark with this signal for providing a watermarked host modified signal m_w, which is also referred to as a first host modifying signal, step 56. The first host modifying signal m_w is then supplied to a second signal shaping unit 44, which uses a second signal shaping function M2 determined by the filter control unit 38 to provide a shaped host modifying signal m_{ws} or second host modifying signal, step 58. The second signal shaping unit 44 is also provided in the form of a filter, the coefficients of which are set according to the above described model P. The function M2 makes sure that there are no extra perceptible artefacts in the watermarked signal. The second host modifying signal m_{ws} is then provided to the second adding unit 36, which also receives the modified host signal $x + n_s$ and adds these two together for providing the watermarked host signal or the watermarked output media signal y, step 60. In this way the watermark is perceptibly masked

by the media signal x. It should be realised that since the noise signal n_s is imperceptibly added to the media signal, it provides an imperceptible watermark channel.

It is possible to vary the function used. As an alternative a so-called threshold-in-quite (TQ) function can be used when the media signal is an audio signal instead of the functions M1 and/or M2 above. In this case the noise is pre-filtered such that it falls below the hearing threshold. Similar functions can be used for image signals and/or video.

The device and method according to the third embodiment of the invention shown in fig. 5 and 6 has a slight disadvantage, which is that the noise signal is added twice to the host signal. This makes the control of the watermarking process slightly unpredictable. A device for the solution of this problem is shown in a block schematic in fig. 7, in a fourth embodiment of the invention. There is no first signal shaping unit in this device. Here the noise signal n is first provided to a scaling unit 62, which scales the noise signal with a scaling function δ. δ is here smaller than one and preferably between 0.1 and 0.2. The downscaled noise signal δn is then supplied to the first adding unit 12 where it is added to the host signal x in order to provide the modified host signal, which is now denoted x + δn because the noise signal has been downscaled. The modified host signal is then passed to the combiner unit 14, which embeds the watermark w in the previously described fashion. The output of the combiner unit 14 is connected to a third adding unit 64, which also receives the unscaled noise signal n for adding to the watermarked modified host signal in order to provide a first host modifying signal m_w. The signal m_w is provided to the second signal shaping unit 44, which filters the first host modifying signal m_w according to the previously described function M2, which is based on the function P of the human hearing system analysis made in filter control unit 38. The shaped signal m_{ws} or second host modifying signal from the filter 44 is provided to the second adding unit 36 for addition to the original host signal x. The filter 44 thus makes sure that the host modifying signal m_w is perceptibly masked by the host signal x. In this way all additional signal components are only injected into the host signal x in one point, which makes the control mechanism more predictable.

As mentioned above the noise signal is added for enabling safer detection of the watermark when the host or media signal has few frequency components, which can be sound frequency components when the signal is an audio signal or spatial frequency components when the signal is an image signal. An audio signal is however not often only made up of spectrally sparse sounds, but can often have few frequency components in just some passages or sections of a piece of music. There can therefore be no need for using the above-described embodiments of the invention in a whole media signal, but only in some

pieces or sections of it. There is thus a need for being able to embed a watermark according to the above-described embodiments of the invention as well as to be able to embed a watermark according to known principles depending on the properties of the media signal.

Fig. 8 shows a device for providing this functionality. The device includes a first adding unit 12, a watermark combiner unit 14 and a second adding unit 36 according to the first embodiment. It should also be realised that the devices according to the second, third and fourth embodiments can easily be adapted to be used in the device in fig. 8 with some slight and straightforward modifications. In fig. 8 the first adding unit 12 receives a noise signal n and a host signal x and adds these together for forming a modified host signal x + n according to the above described principles. The output of the first adding unit 12 is connected to the watermark combiner unit 14 via a first switch 68. The host signal is also directly connected to the watermark combiner unit 14 via a second switch 70. An analysing unit 66 uses an analysing function A for analysing the frequency content of the host signal and controls the first and the second switch in dependence of the analysis, such that the first switch 68 connects the first adding unit 12 to the watermark combiner unit 14 if the number of frequency components in the host signal x are sparse and otherwise the second switch 70 connects the unmodified host signal x to the watermark combiner unit 14. The watermark combiner unit 14 then embeds the watermark in the signal it receives in the previously described fashion, and the second adding unit 36 adds the first host modifying signal m_w to the unmodified host signal x or modified host signal x + n for provision of the output signal y. Here the switching is preferably a soft switching function so that the transition from inputting of one signal to the watermark combiner unit 14 to the other is made gracefully. This means that when switching is performed from one state to another, the switch that is switched on is gradually made to let the signal pass through such that at first it is very small or attenuated and gradually rises until the full signal is being passed through the switch. The switch, which is switching off, is in the same way gradually attenuating the signal it is to switch off until it is completely switched off. This is also preferably done so that the total energy passed through to the watermark combiner unit is substantially unitary both before, during and after switching.

It should be realised that the switching does not have to be soft or graceful, although this is preferred. In case no soft switching is performed, it might be sufficient to only provide one switch, which either connects the modified host signal, or the unmodified host signal to the watermark combiner unit 14. When a single switch is used it is furthermore

possible to provide it in any position which achieves the proper switching of the signals, like for instance before the first adding unit 12.

The output signal y can be provided on a storage medium, of which one 72 in the form of a CD disc is shown in fig. 9. The output signal y can also be provided on other types of storage mediums, such as memory in a computer.

There has thus been described a device and a method for multiplicatively embedding additional data in a media signal when the media signal has few frequency components. With the invention it is possible to embed a watermark in such a media signal which is easier to detect than an ordinary media signal having these properties. The second embodiment makes sure that the added noise is not perceptible and the third embodiment makes sure that both the added noise and the embedded watermark are not perceptible. The fourth embodiment has the advantage of providing a more predictable control mechanism for the embedding of a watermark. A higher level of detectability has furthermore the following advantages. The additional data remains detectable even if the quality of the media signal is degraded. It is then easier to perform for instance copy control or forensic tracking of a processed media signal.

The invention can be varied in many ways. It is for instance possible that the noise signal can be made to include data. This can be made in the way that one random sequence can be made to represent a "zero" and another can be made to represent a "one". In this way additive and multiplicative watermarks can be integrated into a single system. As mentioned before the watermark can be embedded in both the time as well as the frequency domain and the media signal can be any type of media signal. A media signal can furthermore be an audio, video or image signal. In the case of audio it can be uncompressed audio such as PCM. The invention is however also possible to apply on compressed media, which in the case of audio can be a MP3 bitstream. However, then the noise has to be appropriately converted to the bitstream. Therefore the present invention is only to be limited by the following claims.